Should we care about linguistics?

Ellie Pavlick
Department of Computer Science
Brown University
Deep Learning is Taking Over NLP!
Deep Learning is Taking Over NLP!

Titles of ACL Papers, 1979
Deep Learning is Taking Over NLP!

Titles of ACL Papers, 1987
Deep Learning is Taking Over NLP!

Titles of ACL Papers, everything pre-2015
Deep Learning is Taking Over NLP!

Titles of ACL Papers, 2017
SOTA on Classic NLP Tasks

Language Modeling
Bengio et al. (2003)

Perplexity

Sentiment Analysis
Socher et al. (2013)

Accuracy

Dependency Parsing
Chen and Manning (2014)

Unlabelled Attachment Score

Machine Translation
Devlin et al. (2014)

BLEU (Ar-En)

<table>
<thead>
<tr>
<th>Task</th>
<th>Method 1</th>
<th>Method 2</th>
<th>Method 3</th>
<th>Method 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Language Modeling</td>
<td>312</td>
<td>252</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sentiment Analysis</td>
<td>46</td>
<td>45.7</td>
<td>41.9</td>
<td></td>
</tr>
<tr>
<td>Dependency Parsing</td>
<td>93</td>
<td>91.8</td>
<td>90.7</td>
<td></td>
</tr>
<tr>
<td>Machine Translation</td>
<td>53</td>
<td>52.8</td>
<td>49.5</td>
<td></td>
</tr>
</tbody>
</table>

*Method 1: Best N-gram
Method 2: Best MLP
Method 3: Naive Bayes
Method 4: RNN
Method 5: Graph-Based Model
Method 6: Neural Model
Method 7: Best Phrase-Based
Method 8: Best Neural
New Enthusiasm for End-to-End NLU Tasks
New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)
New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

A man inspects the uniform of a figure in some East Asian country.

+ The man is sleeping.
New Enthusiasm for End-to-End NLU Tasks
Recognizing Textual Entailment (RTE)

premise

A man inspects the uniform of a figure in some East Asian country.

+ The man is sleeping.
New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

Premise: A man inspects the uniform of a figure in some East Asian country.

Hypothesis: The man is sleeping.
New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

A man inspects the uniform of a figure in some East Asian country.

+ The man is sleeping.

System
New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

A man inspects the uniform of a figure in some East Asian country.

The man is sleeping.

System

False
New Enthusiasm for End-to-End NLU Tasks

Recognizing Textual Entailment (RTE)

Performance of Sentence Encoding Models on SNLI Dataset

Models on SNLI Leaderboard

A large annotated corpus for learning natural language inference.
Bowman et al. (2015)
New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

What is Southern California often abbreviated as?
Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.
Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.
Reading Comprehension

What is Southern California often abbreviated as?

Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.
New Enthusiasm for End-to-End NLU Tasks

Reading Comprehension

Performance on SQUAD Reading Comprehension Dataset

SQuAD: 100,000+ Questions for Machine Comprehension of Text. Rajpurkar et al. (2016)
Should we care about linguistics?
As is, we are doing lots of tasks very well.
What are our systems learning?

A man inspects the uniform of a figure in some East Asian country.

The man is sleeping.

SNLI dataset (Bowman, 2015)
What are our systems learning?

A man *inspects* the uniform of a figure in some East Asian country.

The man is *sleeping*.

SNLI dataset (Bowman, 2015)
What are our systems learning?

A man inspects the uniform of a figure in some East Asian country.

The man is sleeping.

SNLI dataset (Bowman, 2015)
A man inspects the uniform of a figure in some East Asian country. The man is sleeping.

What are our systems learning?

SNLI dataset (Bowman, 2015)
What are our systems learning?

- Contradiction: 83%
- Entailment: 10%
- Neutral: 7%

Current SOTA is 86%

Tao Shen et al. 2018
300D Reinforced Self-Attention Network

SNLI dataset (Bowman, 2015)

A man is sleeping.
Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.
Southern California, often abbreviated SoCal, is a geographic and cultural region that generally comprises California's southernmost 10 counties. The region is traditionally described as "eight counties", based on demographics and economic ties: Imperial, Los Angeles, Orange, Riverside, San Bernardino, San Diego, Santa Barbara, and Ventura. The more extensive 10-county definition, including Kern and San Luis Obispo counties, is also used based on historical political divisions. Southern California is a major economic center for the state of California and the United States.

What is Southern California often abbreviated as?

Norther California is often abbreviated NorCal.

SoCal

Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)
What are our systems learning?

Accuracy on SQUAD (Reading Comprehension) before and after inserting adversarial examples

Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)
What are our systems learning?

Accuracy on SQUAD (Reading Comprehension) before and after inserting adversarial examples

- ReasoNet-E
- SEDT-E
- BiDAF-E
- Mnemonic-E
- Ruminating
- jNet

Adversarial Examples for Evaluating Reading Comprehension Systems
Jia and Liang (2017)
What do we want our systems to learn?
What do we want our systems to learn?

Pre-Trained Word Embeddings

Word Embedding-based Antonym Detection using Thesauri and Distributional Information (Ono et al. 2015)

Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints (Liu et al. 2015)

SENSEMBED: Learning Sense Embeddings… (Iacobacci et al 2015)

AutoExtend: Extending Word Embeddings…(Rothe and Schutze 2015)

Low-Dimensional Embeddings of Logic (Rocktaschel et al. 2014)

Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction (Nguyen et al. 2016)

Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment (Roller and Erk 2016)

Counter-fitting Word Vectors to Linguistic Constraints (Mrkšić et al. 2016)

Retrofitting Word Vectors to Semantic Lexicons (Faruqui et al 2015)
What do we want our systems to learn?

**Pre-Trained Word Embeddings**
- Low-Dimensional Embeddings of Logic (Rocktaschel et al. 2014)
- Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints (Liu et al. 2015)
- SENSEMBED: Learning Sense Embeddings... (Iacobacci et al. 2015)
- AutoExtend: Extending Word Embeddings... (Rothe and Schutze 2015)

**Task-Independent Sentence Embeddings**
- Supervised Learning of Universal Sentence Representations from Natural Language Inference Data (Conneau et al. 2017)
- Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection (Socher et al. 2011)

**Retrofitting Word Vectors to Semantic Lexicons** (Faruqui et al 2015)

**Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction** (Nguyen et al. 2016)

**Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment** (Roller and Erk 2016)

**Counter-fitting Word Vectors to Linguistic Constraints** (Mrkšić et al. 2016)

**A Structured Self-attentive Sentence Embedding** (Lin et al. 2017)

**Siamese CBOW: Optimizing Word Embeddings for Sentence Representations** (Kenter et al. 2016)

**Towards Universal Paraphrastic Sentence Embeddings** (Wieting et al. 2016)
What do we want our systems to learn?

Word Embedding-based Antonym Detection using Thesauri and Distributional Information (Ono et al. 2015)

Pre-Trained Word Embeddings

Retrofitting Word Vectors to Semantic Lexicons (Faruqui et al. 2015)

Low-Dimensional Embeddings of Logic (Rocktaschel et al. 2014)

Counter-fitting Word Vectors to Linguistic Constraints (Mrkšić et al. 2016)

Learning Semantic Word Embeddings based on Ordinal Knowledge Constraints (Liu et al. 2015)

Integrating Distributional Lexical Contrast into Word Embeddings for Antonym-Synonym Distinction (Nguyen et al. 2016)

SENSEMBED: Learning Sense Embeddings… (Iacobacci et al 2015)

Identifying and Exploiting Hearst Patterns in Distributional Vectors for Lexical Entailment (Roller and Erk 2016)

AutoExtend: Extending Word Embeddings…(Rothe and Schutze 2015)

Skip-Thought Vectors (Kiros et al. 2015)

Distributed Representations of Sentences and Documents (Quoc et al. 2014)

A Structured Self-attentive Sentence Embedding (Lin et al. 2017)

Supervised Learning of Universal Sentence Representations from Natural Language Inference Data (Conneau et al. 2017)

Task-Independent Sentence Embeddings

Towards Universal Paraphrastic Sentence Embeddings (Wieting et al. 2016)

Dynamic Pooling and Unfolding Recursive Autoencoders for Paraphrase Detection (Socher et al. 2011)

Deep Visual-Semantic Alignments for Generating Image Descriptions (Karpathy and Fei Fei 2015)

Knowledge Representation and Grounding

Improved Representation Learning for Predicting Commonsense Ontologies (Li et al 2017)

Deep Visual-Semantic Alignments for Generating Image Descriptions (Karpathy and Fei Fei 2015)

Embedding Multimodal Relational Data (Open review 2018)

Learning Structured Embeddings of Knowledge Bases (Bordes et al. 2011)

Multimodal Neural Language Models (Kiros et al 2014)

DeViSE: A Deep Visual-Semantic Embedding Model (Frome et al 2013)

Learning Semantic Hierarchies via Word Embeddings (Fu et al. 2014)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them…

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks…

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them…

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
This workshop deals with the evaluation of general-purpose vector representations for linguistic units (morphemes, words, phrases, sentences, etc). What distinguishes these representations (or embeddings) is that they are not trained with a specific application in mind, but rather to capture broadly useful features of the represented units. Another way to view their usage is through the lens of transfer learning: The embeddings are trained with one objective, but applied on others.

Evaluating general-purpose representation learning systems is fundamentally difficult. They can be trained on a variety of objectives, making simple intrinsic evaluations useless as a means of comparing methods. They are also meant to be applied to a variety of downstream tasks, which will place different demands on them…

RepEval 2017
(Bowman, Goldberg, Hill, Lazaridou, Levy, Reichart, and Søgaard)
“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.”

—Richard Montague
“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.”

—Richard Montague
“There is in my opinion no important theoretical difference between natural languages and the artificial languages of logicians; indeed I consider it possible to comprehend the syntax and semantics of both kinds of languages with a single natural and mathematically precise theory.”

—Richard Montague
Language → Math
Language → ∀x∀y(P(f(x)) → ¬(Q(f(y),x)))
Language $\rightarrow \lambda g. (\lambda x. g \ (x \ x)) (\lambda x. g \ (x \ x))$
Language $\rightarrow c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$

$h_t = o_t \circ \sigma_h(c_t)$
Language → 0101101101010010101010
Language → 0101101101010010101010
The Semantics-Pragmatics Interface
The Semantics-Pragmatics Interface
The Semantics-Pragmatics Interface

I went to the beach over vacation.
The Semantics-Pragmatics Interface

I went to the beach over vacation.
The Semantics-Pragmatics Interface

I went to the beach over vacation.
The Semantics-Pragmatics Interface

I went to the beach over vacation.
The Semantics-Pragmatics Interface

I went to the beach over vacation.

Context TBD
The Semantics-Pragmatics Interface

I went to the beach over vacation.

I laid out in the sun.

Context TBD
The Semantics-Pragmatics Interface

I went to the beach over vacation.

I laid out in the sun.
What “belongs” in the representation of a word?

beach
What “belongs” in the representation of a word?

location near the water

beach
What “belongs” in the representation of a word?

location near the water

beach

is a place
What “belongs” in the representation of a word?

location near the water

beach

is a place

is a popular vacation place
What “belongs” in the representation of a word?

- location near the water
- beach
- is a place
- not indoors
- is a popular vacation place
What “belongs” in the representation of a word?

- location near the water
- is a place
- not indoors
- is a popular vacation place
- may have palm trees
What “belongs” in the representation of a word?

location near the water

beach

is a place

not indoors

is a popular vacation place

may have palm trees

P(palm trees)
What “belongs” in the representation of a word?

- location near the water
- beach
- is a place
- not indoors
- is a popular vacation place
- P(it is okay for a listener to imagine that it has palm trees)
- P(palm trees)
- may have palm trees
What “belongs” in the representation of a word?

location near the water

beach

is a place not indoors may have palm trees

P(palm trees)

P(it is okay for a listener to imagine that it has palm trees)

Dictionary Representation

is a popular vacation place
What “belongs” in the representation of a word?

- location near the water
- is a place
- not indoors
- is a popular vacation place
- P(it is okay for a listener to imagine that it has palm trees)
- P(palm trees)

Taxonomic Representation
What “belongs” in the representation of a word?

beach

Knowledge Base Representation

P(it is okay for a listener to imagine that it has palm trees)

location near the water

is a place

not indoors

may have palm trees

is a popular vacation place
What “belongs” in the representation of a word?

P(it is okay for a listener to imagine that it has palm trees)

Prototype Representation

beach

location near the water

is a place

not indoors

is a popular vacation place

P(palm trees)
What “belongs” in the representation of a word?

location near the water

beach

is a place

is a popular vacation place

“Encyclopedic” Representation

P(it is okay for a listener to imagine that it has palm trees)

P(palm trees)

may have palm trees

not indoors
Is SkipGram enough?

P(occurs after “on”)
P(occurs after “the”)
P(occurs before “vacation”)
P(occurs after “clandestine”)
P(occurs before “grinning”)

Distributional Contextual Representation
Should we care about linguistics?
Should we care about linguistics?

Yes.
Should we care about linguistics?

Yes.
Because we have to form and test hypotheses about what our word representations should capture.
What “belongs” in the representation of a word?

beach

Model Theoretic Representation
What “belongs” in the representation of a word?

beach
What “belongs” in the representation of a word?
What “belongs” in the representation of a word?
What “belongs” in the representation of a word?

beach

sandy beach

tropical beach
What “belongs” in the representation of a word?
Set-Theoretic Semantics

A little boy doing a hand stand on the beach.
Set-Theoretic Semantics

A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.
Set-Theoretic Semantics

$$\exists x ( \text{beach}(x) \land \neg \text{sandy\_beach}(x) )$$

A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.
Set-Theoretic Semantics

\[ \exists x ( \text{beach}(x) \land \neg \text{sandy_beach}(x) ) \]

Set-theoretic semantics does not allow this inference.

A little boy doing a hand stand on the beach.

\[ \downarrow \]

A little boy doing a hand stand on the sandy beach.
Set-Theoretic Semantics?

A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.

Most “babies” are “little” and most “problems” are “huge”: Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Set-Theoretic Semantics?

A little boy doing a hand stand on the beach.

+ 

A little boy doing a hand stand on the sandy beach.

\[ \text{Human Annotators} \]

Most “babies” are “little” and most “problems” are “huge”: Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)
Set-Theoretic Semantics?

A little boy doing a hand stand on the beach. +

A little boy doing a hand stand on the sandy beach.

Human Annotators

Yes

Most “babies” are “little” and most “problems” are “huge”:

Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Set-Theoretic Semantics?

\[ \exists x ( \text{beach}(x) \land \neg \text{sandy\_beach}(x) ) \]

A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.

Set-theoretic semantics does not allow this inference.
Set-Theoretic Semantics?

$\exists x (\text{beach}(x) \land \neg \text{sandy\_beach}(x))$

A little boy doing a hand stand on the beach.

A little boy doing a hand stand on the sandy beach.

Set-theoretic semantics does not allow this inference.

Human subjects unanimously agree that this inference is valid.
Set-Theoretic Semantics?

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Set-Theoretic Semantics?

Most “babies” are “little” and most “problems” are “huge”:

Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)
Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
In the worst case, Model-Theoretic representation makes incorrect predictions 47% of the time!

Set-Theoretic Semantics?

Images

News

Literature

Debate Forums

Most "babies" are "little" and most "problems" are "huge": Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)
### Human Inferences

<table>
<thead>
<tr>
<th>P entails H</th>
<th>P contradicts H</th>
</tr>
</thead>
<tbody>
<tr>
<td>Somehow, I feel there will be a lack of evidence forthcoming</td>
<td>Bush travels Monday to Michigan to make remarks on the economy.</td>
</tr>
<tr>
<td><strong>evidence</strong> -&gt; <strong>credible evidence</strong></td>
<td><strong>economy</strong> -&gt; <strong>Japanese economy</strong></td>
</tr>
<tr>
<td>Penfield Evans grasped his <strong>hand</strong> and shook it warmly.</td>
<td>Government is the only thing holding back large corporations.</td>
</tr>
<tr>
<td><strong>hand</strong> -&gt; <strong>outstretched hand</strong></td>
<td><strong>government</strong> -&gt; <strong>small government</strong></td>
</tr>
<tr>
<td>His <strong>body</strong> is found a week later.</td>
<td>A child rides on a <strong>man’s shoulders</strong>.</td>
</tr>
<tr>
<td><strong>body</strong> -&gt; <strong>dead body</strong></td>
<td><strong>man</strong> -&gt; <strong>homeless man</strong></td>
</tr>
</tbody>
</table>

Pavlick and Callison-Burch (2016)
What “belongs” in the representation of a word?

- evidence -> credible evidence
- economy -> Japanese economy
- hand -> outstretched hand
- government -> small government
- body -> dead body
- man -> homeless man
What “belongs” in the representation of a word?

- evidence -> is credible?
- body -> is dead?
- government -> isn’t small?
- man -> isn’t homeless?
- hand -> is outstretched?
- economy -> isn’t Japanese?
What “belongs” in the representation of a word?

**Semantics**
- evidence -> is credible?
- body -> is dead?
- government -> isn’t small?
- man -> isn’t homeless?

**Pragmatics**
- hand -> is outstretched?
- economy -> isn’t Japanese?
What “belongs” in the representation of a word?

**Semantics**
- evidence -> is credible?
- body -> is dead?
- government -> isn’t small?

**Pragmatics**
- man -> isn’t homeless?
- hand -> is outstretched?
- economy -> isn’t Japanese?
What “belongs” in the representation of a word?

Semantics

- evidence -> is credible?
- body -> is dead?
- government -> isn’t small?

Pragmatics

- man -> isn’t homeless?
- hand -> is outstretched?
- economy -> isn’t Japanese?
evidence -> is credible?
  body -> is dead?
  government -> isn’t small?
  man -> isn’t homeless?
  hand -> is outstretched?
  economy -> isn’t Japanese?
Should we care about linguistics?
Recognizing Textual Entailment Task

Most “babies” are “little” and most “problems” are “huge”: Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Recognizing Textual Entailment Task

A group of hikers walk a path that leads from a sandy beach towards a hill

+ The hikers are walking outside

RTE System

True

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Simplified RTE Task

A hiker walking on a path at the foot of snow capped mountains

+ A hiker walking on a sandy path at the foot of snow capped mountains

RTE System

False

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Simplified RTE Task

- 5,378 add-one pairs

Most “babies” are “little” and most “problems” are “huge”: Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Simplified RTE Task

- 5,378 add-one pairs
- 4,991 for training (4,481 training, 510 dev)

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Simplified RTE Task

• 5,378 add-one pairs
• 4,991 for training (4,481 training, 510 dev)
• 387 test (removed pairs with low human agreement)

Most “babies” are “little” and most “problems” are “huge”:
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Simplified RTE Task

- 5,378 add-one pairs
- 4,991 for training (4,481 training, 510 dev)
- 387 test (removed pairs with low human agreement)
- 500K general RTE pairs from SNLI

Most “babies” are “little” and most “problems” are “huge”: Compositional Entailment in Adjective-Nouns. Pavlick and Callison-Burch (2016)
Simplified RTE Task

<table>
<thead>
<tr>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
</tr>
<tr>
<td>92</td>
</tr>
<tr>
<td>84</td>
</tr>
<tr>
<td>76</td>
</tr>
<tr>
<td>68</td>
</tr>
<tr>
<td>60</td>
</tr>
</tbody>
</table>
Simplified RTE Task

Always Predict "Non-Entailment"

Accuracy

85.3

Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Most "babies" are "little" and most "problems" are "huge":

Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)

Accuracy

Always Predict "Non-Entailment"

92.2

Most Frequent Class by Adjective

Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)
Most "babies" are "little" and most "problems" are "huge":

Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

Simplified RTE Task

Accuracy

85.3
86
92.2

Always Predict "Non-Entailment"
Bag of Words
Most Frequent Class by Adjective

Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Most "babies" are "little" and most "problems" are "huge":

Compositional Entailment in Adjective-Nouns

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always Predict &quot;Non-Entailment&quot;</td>
<td>85.3</td>
</tr>
<tr>
<td>Bag of Words</td>
<td>86</td>
</tr>
<tr>
<td>Bag of Vectors</td>
<td>86.6</td>
</tr>
<tr>
<td>Most Frequent Class by Adjective</td>
<td>92.2</td>
</tr>
</tbody>
</table>

Simplified RTE Task

Pavlick and Callison-Burch (2016)
Most "babies" are "little" and most "problems" are "huge": Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)
Most "babies" are "little" and most "problems" are "huge":

Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always Predict &quot;Non-Entailment&quot;</td>
<td>85.3</td>
</tr>
<tr>
<td>Bag of Words</td>
<td>86</td>
</tr>
<tr>
<td>Bag of Vectors</td>
<td>86.6</td>
</tr>
<tr>
<td>RNN</td>
<td>87.3</td>
</tr>
<tr>
<td>LSTM</td>
<td>86.6</td>
</tr>
<tr>
<td>Most Frequent Class by Adjective</td>
<td>92.2</td>
</tr>
</tbody>
</table>
Most "babies" are "little" and most "problems" are "huge":
Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always Predict &quot;Non-Entailment&quot;</td>
<td>85.3</td>
</tr>
<tr>
<td>Bag of Words</td>
<td>86</td>
</tr>
<tr>
<td>Bag of Vectors</td>
<td>86.6</td>
</tr>
<tr>
<td>RNN</td>
<td>87.3</td>
</tr>
<tr>
<td>LSTM</td>
<td>86.6</td>
</tr>
<tr>
<td>LSTM + Transfer Bowman et al. (2015)</td>
<td>86.8</td>
</tr>
<tr>
<td>Most Frequent Class by Adjective</td>
<td>92.2</td>
</tr>
</tbody>
</table>

Simplified RTE Task

Compositional Entailment in Adjective-Nouns
Pavlick and Callison-Burch (2016)
Most "babies" are "little" and most "problems" are "huge":

Compositional Entailment in Adjective-Nouns

Pavlick and Callison-Burch (2016)

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Always Predict &quot;Non-Entailment&quot;</td>
<td>85.3</td>
</tr>
<tr>
<td>Bag of Words</td>
<td>86</td>
</tr>
<tr>
<td>Bag of Vectors</td>
<td>86.6</td>
</tr>
<tr>
<td>RNN</td>
<td>87.3</td>
</tr>
<tr>
<td>LSTM</td>
<td>86.6</td>
</tr>
<tr>
<td>LSTM + Transfer Bowman et al. (2015)</td>
<td>86.8</td>
</tr>
</tbody>
</table>

Most Frequent Class by Adjective

Simplified RTE Task
Simplified RTE Task

Add-One Adjective

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM Add-One data only</td>
<td>86.6</td>
</tr>
<tr>
<td>LSTM + Transfer</td>
<td>86.8</td>
</tr>
</tbody>
</table>

LSTM + Transfer

SICK

Accuracy

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>LSTM SICK data only</td>
<td>71.3</td>
</tr>
<tr>
<td>LSTM + Transfer</td>
<td>80.8</td>
</tr>
</tbody>
</table>
Should we care about linguistics? Yes!

Because we want to learn task-independent representations of language, which requires asking and answering:

1. What components of linguistic meaning are "intrinsic", and what is derived in context/at "runtime"?

2. If these representation can't be trained in end-to-end tasks: how to we know what is the "right" representation? Which tasks should be viewed as "fundamental" and trained/test explicitly, and which ones should come along "for free"?

Takeaways
Takeaways

• Should we care about linguistics? Yes!
Takeaways

- Should we care about linguistics? Yes!

- Because we want to learn task-independent representations of language, which requires asking and answering:
Takeaways

• Should we care about linguistics? Yes!

• Because we want to learn task-independent representations of language, which requires asking and answering:

  1. What components of linguistic meaning are “intrinsic”, and what is derived in context/at “runtime”?
Takeaways

• Should we care about linguistics? Yes!

• Because we want to learn task-independent representations of language, which requires asking and answering:

  1. What components of linguistic meaning are “intrinsic”, and what is derived in context/at “runtime”?

  2. If these representation can’t be trained in end-to-end tasks: how to we know what is the “right” representation? Which tasks should be viewed as “fundamental” and trained/test explicitly, and which ones should come along “for free”?
Thank you!